Molecular Phylogenetics and Evolution 61 (2011) 157–166

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Bayesian relaxed clock estimation of divergence times in foraminifera

Mathieu Groussin [a,1], Jan Pawlowski [b], Ziheng Yang [a,c,*]

[a] Galton Laboratory, Department of Biology, University College London, London WC1E 6BT, England, United Kingdom
[b] Department of Genetics and Evolution, University of Geneva, CH-1211 Geneva, Switzerland
[c] Centre for Computational and Evolutionary Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Article history:
Received 13 February 2011
Revised 17 May 2011
Accepted 10 June 2011
Available online 23 June 2011

Keywords:
Foraminifera
Divergence times
Bayesian method
Relaxed clock
MCMC
Infinite-sites plot

ABSTRACT

Accurate and precise estimation of divergence times during the Neo-Proterozoic is necessary to understand the speciation dynamic of early Eukaryotes. However such deep divergences are difficult to date, as the molecular clock is seriously violated. Recent improvements in Bayesian molecular dating techniques allow the relaxation of the molecular clock hypothesis as well as incorporation of multiple and flexible fossil calibrations. Divergence times can then be estimated even when the evolutionary rate varies among lineages and even when the fossil calibrations involve substantial uncertainties. In this paper, we used a Bayesian method to estimate divergence times in Foraminifera, a group of unicellular eukaryotes, known for their excellent fossil record but also for the high evolutionary rates of their genomes. Based on multigene data we reconstructed the phylogeny of Foraminifera and dated their origin and the major radiation events. Our estimates suggest that Foraminifera emerged during the Cryogenian (650–920 Ma, Neo-Proterozoic), with a mean time around 770 Ma, about 220 Myr before the first appearance of reliable foraminiferal fossils in sediments (545 Ma). Most dates are in agreement with the fossil record, but in general our results suggest earlier origins of foraminiferal orders. We found that the posterior time estimates were robust to specifications of the prior. Our results highlight inter-species variations of evolutionary rates in Foraminifera. Their effect was partially overcome by using the partitioned Bayesian analysis to accommodate rate heterogeneity among data partitions and using the relaxed molecular clock to account for changing evolutionary rates. However, more coding genes appear necessary to obtain more precise estimates of divergence times and to resolve the conflicts between fossil and molecular date estimates.

1. Introduction

Foraminifera belong to the eukaryotic super-group Rhizaria and are arguably one of the most important protist groups on Earth (Keeling et al., 2005). They are widely known for inhabiting marine ecosystems, but occupy freshwater and terrestrial environments as well (Holzmann et al., 2003; Lejzerowicz et al., 2010). In ocean habitats, foraminifers can have benthic or planktonic mode of life. The phylogenetic relationship between benthic and planktonic species are currently controversial (Ujiié et al., 2008) but planktonic species are known to appear later in the fossil record. A hallmark of foraminifers is their shells, more specifically termed "tests", which can either be organic, agglutinated or calcareous. Further variations in test morphology exist between taxa due to the construction of structurally distinct unilocular (single chamber) and multilocular (multi-chambers) tests. In general, the organic walls, which are present only in unilocular taxa, are thin and consist of an association between proteins and mucopolysaccharides. The agglutinated foraminifers form their test by cementing environmental particles and may have one or several chambers. Finally, calcareous foraminifers secrete calcium carbonate, principally as calcite, to constitute the wall of single- or multilocular tests.

Foraminifera possess one of the most profuse fossil records among eukaryotes. The earliest Cambrian foraminiferal genus *Platysolenites* has the appearance of a large, simple, agglutinated tube resembling modern foraminiferal genus *Bathysiphon* (McIlroy et al., 2001). Other straight and coiled tubular agglutinated foraminifera, including genus *Ammodiscus* have been reported from the Lower and Middle Cambrian (Culver, 1991). Some studies mention the possibility that unilocular agglutinated foraminifers were already present during the Upper Vendian (Ediacaran period), at the end of the Neo-Proterozoic (Gaucher and Sprechmann, 1999). However, this proposal is controversial given the difficulty in attributing these fossils to Foraminifera with confidence. A recent

* Corresponding author. Address: Department of Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, England, United Kingdom. Fax: +44 (20) 7679 7096.
E-mail address: z.yang@ucl.ac.uk (Z. Yang).
[1] Present address: Université de Lyon, F-69000, Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622 Villeurbanne, France.

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.06.008

study (Pawlowski et al., 2003) proposed that Foraminifera actually emerged during the Neo-Proterozoic, but was unable to provide a more concise time interval for the emergence than between 690 and 1150 Ma. However, this approach, based on molecular divergence time estimates, may have been limited by the exclusive use of partial ribosomal SSU sequences.

Until recently, molecular phylogeny of Foraminifera was inferred almost solely from ribosomal DNA sequences (Pawlowski et al., 1994, 1997, 1999, 2003; Ertan et al., 2004; Schweizer et al., 2008), with only a few phylogenetic analyses of protein coding genes (Habura et al., 2006; Longet and Pawlowski, 2007). For the majority of foraminiferal species, the only sequences available are the partial SSU rDNA sequences, characterized by variable substitution rates and unequal sequence lengths due to numerous insertion events (Pawlowski et al., 1997; Pawlowski and Holzmann, 2002). Consequently, the SSU rDNA alignments suffer from intense removal of sites, which leads to phylogenetic trees with weak resolutions. In spite of these drawbacks, the general view of foraminiferal phylogeny derived from these alignments is congruent in many respects with paleontological data. According to this view, the basal foraminiferal group is composed of monothalamiid (unilocular) organic-walled or agglutinated species, which gave rise to polythalamous (multilocular) clades at least twice during their history (Pawlowski and Holzmann, 2002). One of the resulting clades groups together agglutinated Textulariida and calcareous Rotaliida, while the other clade groups all lineages with early tubular chambers, including Miliolida, Spirillinida and some Lituolinida (*Ammodiscus*, *Miliammina*) (Pawlowski et al., 2003) (see Fig. 1). The earliest multichambered agglutinated Textulariida arose in the Devonian (>400 Ma), while the calcareous porcellaneous Miliolida are known since Carboniferous (>350 Ma) (Haynes, 1981). Nevertheless, the deep relationships between monothalamous and polythalamous lineages remain unsolved.

The molecular clock hypothesis provides a seductively powerful way to date evolutionary events such as speciation. However, the use of a strict molecular clock can lead to seriously biased estimates of divergence times when the clock is violated. The development of new algorithms in the likelihood and Bayesian frameworks has allowed different lineages to have variable evolutionary rates, thus improving the estimation of divergence times and reconciling palaeontologists and molecular systematists (Hasegawa et al., 2003; Douzery et al., 2004; Bell and Donoghue, 2005).

The well-studied fossil record of Foraminifera provides valuable information for calibrating the molecular phylogeny, to date the emergence and divergence times of major taxa. We should bear in mind, however, that the origin of Foraminifera is undoubtedly prior to their first appearance in the fossil record (545 Ma, Culver, 1991; McIlroy et al., 2001) and that this time difference is key in furthering our understanding of the early dynamics of Eukaryotes. To this end, the Bayesian statistical framework is of great interest because it permits the use of prior knowledge about times and rates. Thus, MCMCTREE in the PAML package (Yang and Rannala, 2006; Rannala and Yang, 2007; Yang, 2007) accommodates the uncertainties present in the fossil record by the use of soft bounds and flexible statistical distributions. In addition, MCMCTREE relaxes the molecular clock hypothesis by implementing two models of variable rates among lineages: the independent-rates model, where the rates for branches are independent variables from the same distribution, and the correlated-rates model, where the evolutionary rate of the daughter branch depends on the rate of the ancestral branch.

In this study, we used complete SSU rDNA sequences and three nuclear gene sequences, in addition to partial SSU rDNA sequences, to infer the phylogeny of Foraminifera and to estimate divergence times of their major lineages. We focused on benthic foraminifers because of the extreme evolutionary distances within planktonic species and between benthic and planktonic species. We conducted an extensive robustness analysis to examine the impact of various prior assumptions on our posterior time estimates. We discuss the conflicts between the molecular and fossil time estimates.

2. Material and methods

2.1. Sequence data

The main data set consists of 34 species representing the known taxonomic diversity of Foraminifera. Whenever possible, complete SSU sequences were extracted from the GenBank. When these are unavailable, partial SSU sequences were used. Three nuclear protein-coding genes were also included: actin-2, β-tubulin and RPB1. The actin-2 and β-tubulin sequences of *Quinqueloculina sp.* were retrieved by a blast procedure among all Expressed Sequence Tags (ESTs) available for this species. To reduce the proportion of missing data in the coding-genes data set, some species belonging to the same genus were merged and referred to by their genus names (*Ammonia*, *Allogromia*, *Bathysiphon*, *Bolivina*, *Reophax* and *Sorites*) (Suppl. Table 1). The SSU rDNA sequence is available for 34 species, while at least one coding-gene sequence is available for only 26 species. See Suppl. Table 1 for details.

2.2. Alignment

All sequences were aligned with Muscle 3.7 (Edgar, 2004). The SSU alignment was then improved by successive re-alignments with Muscle within regions of large insertions located between parts of conserved regions, by using the latest version of SeaView (Gouy et al., 2010). Instead of manually improving the SSU alignment by removing fast-evolving regions, we used the SlowFaster program (Kostka et al., 2008) to do so automatically. This requires a prior topology and uses maximum parsimony to establish different thresholds of evolutionary rates within pre-defined monophyletic groups. Alignments of different sizes were generated, depending on the number of fast sites removed, and for each alignment, the maximum likelihood (ML) tree was reconstructed using PhyML (Guindon and Gascuel, 2003) under the GTR + Γ_5 + I model with 100 bootstrap replicates. We chose for all later analysis of this paper the alignment that showed the highest bootstrap support values for the following nodes: origins of rotaliids and recent miliolids (*Quinqueloculina sp.*, *Pyrgo peruviana*, *Peneroplis pertusus*, *Marginopora vertebralis*, *Sorites* and *Amphisorus hemprichii*) and the origin of the group *Ammonia* + *Haynesina germanica* + *Elphidium williamsoni*. Those nodes were chosen because they were widely accepted to be monophyletic. This heuristic approach was taken mainly because numerous insertions and fast-evolving sites cause difficulties in the alignment. As the topology for the entire foraminiferal domain contains uncertainties, the input topology used by SlowFaster included multifurcations at unresolved nodes and respected the well-known and supported monophyly of rotaliids and recent miliolids. The Slow–Fast approach was not used to obtain the coding-genes alignment, as it is straightforward to obtain reliable alignments.

As a result, several data sets were compiled. SSU rDNA with 1942 sites, and the first and second codon positions of the three coding genes with 2148 sites. For phylogeny reconstruction, the data are analyzed using two partitioning strategies. In the first, the SSU rDNA data (1942 sites) and the first and second codon positions data (2148 sites) were concatenated into one partition, with 4099 sites in total. The second strategy treats the data as two partitions: the SSU rDNA vs. positions 1 + 2 of the coding genes. The dating analysis was conducted using the MCMCTREE

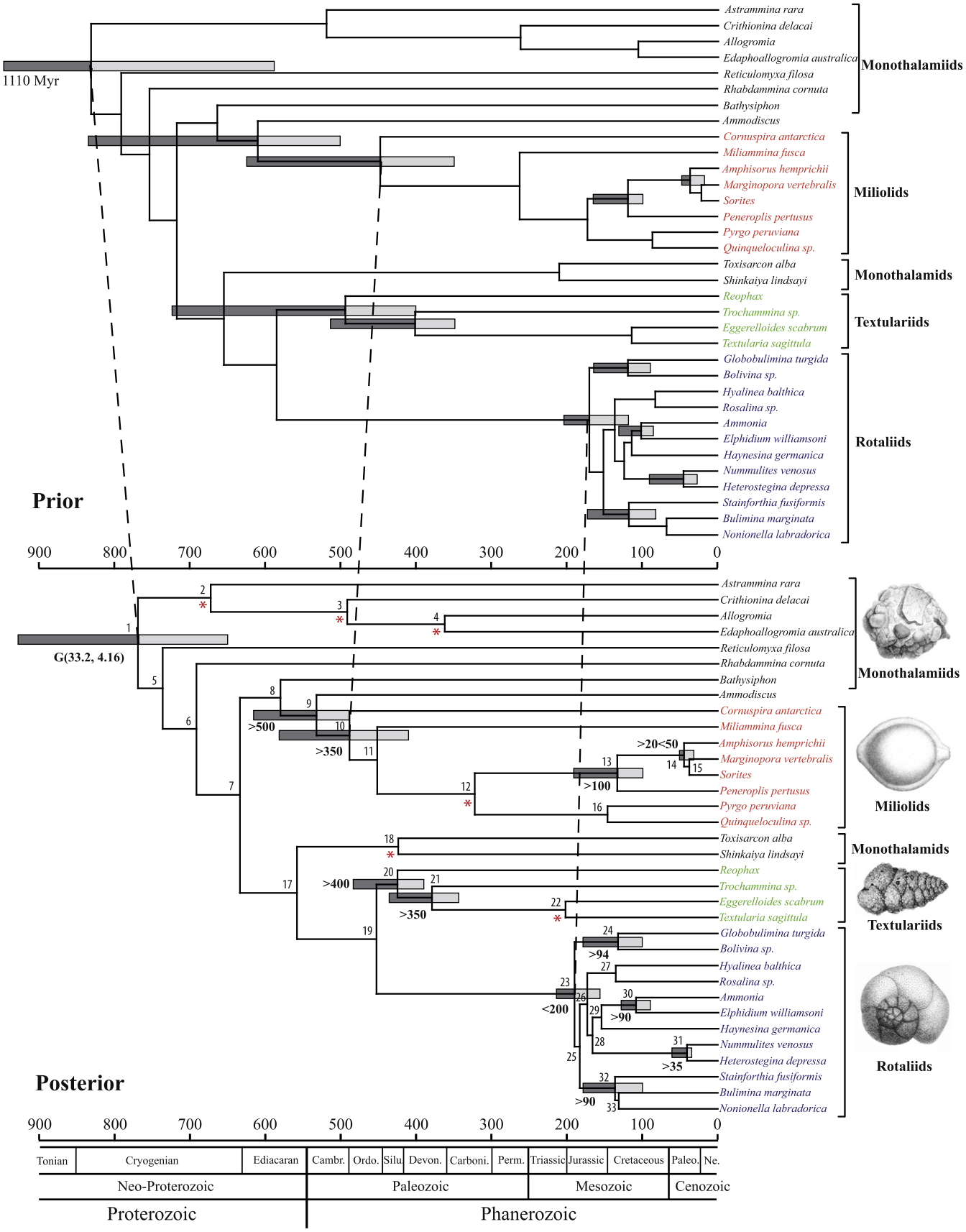

Fig. 1. The phylogeny of Foraminifera showing fossil calibrations and prior (top) and posterior (bottom) means of divergence times. Eleven fossil calibrations are used in the dating analysis, including nine minimum bounds, a maximum bound of 200 Myr for the root of Rotaliids and a pair of joint bounds (>20 < 50) for the origin of Soritinae. A gamma prior *G*(33.2, 4.16) is assigned on the age of the root. The 95% credibility intervals are shown for all nodes where fossil calibrations are available.

program, with the data analyzed as two partitions (rDNA vs. positions 1 + 2 of the coding genes) or four partitions (rDNA vs. three partitions for the three genes). Each partition had its own set of substitution parameters and substitution rates (Yang, 1996). Divergence times were also estimated using two mixed partitions (rDNA vs. amino acids).

2.3. Phylogenetic reconstructions

For the concatenated data set, the ML tree was reconstructed using the PhyML program (Guindon and Gascuel, 2003) under the GTR + Γ_5 + I model, with 100 bootstrap replicates. The Bayesian tree was reconstructed using PhyloBayes (Lartillot et al., 2009) under the GTR + CAT + Γ_5 model. Trace plots of the log-likelihood values confirmed convergence of the MCMC. Two MCMC runs were used to confirm consistency between runs. We also conducted a phylogenetic analysis on the two-partitions data set (see above) with RAxML (Stamatakis, 2006) and MrBayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The GTR + Γ_5 model, with 100 bootstrap replicates was used in RAxML, starting from a random tree. Similarly, the GTR + Γ_5 model was applied with MrBayes.

2.4. Calibration points and Bayesian divergence time estimation

Both the ML and Bayesian analyses produced unrooted trees. These were integrated with previous analyses of molecular and fossil data to generate a rooted tree for Bayesian estimation of divergence times using MCMCTREE.

Twelve fossil calibrations were used to calibrate the foraminiferal tree, implemented using soft bounds (Yang and Rannala, 2006) (Table 1). Nine of them represent minimum bounds and are assigned the truncated Cauchy distribution with $p = 0.1$ and $c = 0.2$ so that the density falls off rapidly away from the mode, which is close to the fossil minimum (Inoue et al., 2010). The microfossil record is well documented, so we expect the true age to be close to this minimum. The fossil calibrations are according to Haynes (1981) and Loeblich and Tappan (1987). A few comments are in order. (i) According to the fossil record, the radiation of Rotaliida occurred in Early Cretaceous but they probably already diverged during the Jurassic period (Haynes, 1981). Thus a maximum bound of 200 Ma was used on this node (node 23 in the tree of Fig. 1). (ii) The first presence of the Soritinae subfamily (*A. hemprichii*, *Marginopora* and *Sorites*) is recorded in rocks since 20 Ma, during Miocene (Haynes, 1981). On this node (node 14), a maximum bound of 50 Ma was applied, based on the fossil apparition of the earliest Archaiasinae, sister group of Soritinae (Holzmann et al., 2001). (iii) The program MCMCTREE requires a constraint on the age of the root in the dating analysis. As no reliable fossil calibration exists for the origin of Foraminifera, a gamma prior was used for the root age. A previous study based on molecular divergence time estimations proposed that Foraminifera could have originated between 690 and 1150 Ma (Pawlowski et al., 2003). The oldest fossils attributed to Foraminifera with confidence are from the early Cambrian (Culver, 1991). So we used the gamma prior $G(33.2, 4.16)$ for the root age, with the mode at 774 Ma and 95% CI to be (550, 1090). Note that the gamma distribution $G(\alpha, \beta)$ has mean α/β and variance α/β^2.

Table 1
Time constraints (Ma) on nodes used during the dating analysis (Fig. 1).

Node number	Calibration	Fossil appearance
1	$G(33.2, 4.16)$	Basal Monothalamids
9	$L(5)$	*Ammodiscus* genus
10	$L(3.5)$	*Cornuspira* genus
13	$L(1)$	Peneroplidae family
14	$B(0.2, 0.5)$	0.2: Soritinae subfamily 0.5: Archaiasinae subfamily (sister group of Soritinae)
20	$L(4)$	*Reophax* genus
21	$L(3.5)$	*Trochammina* genus
23	$U(2)$	Rotaliida order
24	$L(0.94)$	Bolivinidae family
30	$L(0.90)$	Rotaliidae family
31	$L(0.35)$	Nummulitidae family (recent species)
32	$L(0.90)$	*Stainforthia* genus

Note: Node numbers refer to those in Fig. 1. The time unit is 100 Myrs. Four kinds of calibrations are used on the ages of nodes: $G(a, b)$ is the gamma distribution with mean a/b and variance a/b^2; $B(a, b)$ is a pair of joint bounds $a < t < b$, implemented using equation (17) in Yang and Rannala (2006); $L(a)$ is the minimum-age (lower) bound, implemented using equation (26) in Inoue et al. (2010) with $p = 0.1$ and $c = 0.2$; and $U(b)$ is the maximum-age (upper) bound $t < b$, implemented using equation (16) in Yang and Rannala (2006).

The univariate calibration densities are multiplied and then truncated so that ancestral nodes are older than descendent nodes, leading to a joint distribution of ages of nodes with calibration information. This is multiplied by the prior density for ages of nodes without calibration information, specified using the birth-death process with species sampling, with the birth and death rates $\lambda = \mu = 2$ and the sampling fraction $\rho = 0.1$, representing a nearly flat kernel density (Yang and Rannala, 2006). As the truncation mentioned above can cause the effective prior used by the program to differ considerably from the apparent prior specified by the user (Inoue et al., 2010), we ran the program without the sequence data to generate the effective prior to make sure that it is reasonable judged by the fossil evidence.

The likelihood was calculated under the HKY85 + Γ_5 model, using both Felsenstein's (1981) exact algorithm and using the normal approximation (Thorne et al., 1998; Yang, 2006: figure 7.10). The two methods produced similar divergence time estimates. Our main results presented below were obtained using the approximate method. For the exact likelihood calculation, a gamma prior $G(6, 2)$ was assigned to the transition/transversion rate ratio κ and a gamma prior $G(1, 1)$ was used for the gamma shape parameter α of the HKY85 + Γ_5 model.

The time unit is set at 100 Myr. To specify a gamma prior for the overall rate μ, we fix its shape parameter at $\alpha = 1$ (which represents a diffuse prior), and obtain the scale parameter β by fitting a molecular clock to the sequence data using point calibrations to estimate the prior mean ($=\alpha/\beta$). This led to the prior $\mu \sim G(1, 30)$, with the mean rate to be 0.033 per time unit or 3.3×10^{-10} substitutions per site per year. Both the independent-rates and the correlated-rates models were used to accommodate variable rates between branches (Rannala and Yang, 2007). The two models produced similar time estimates (see below) so our main results are presented under the independent-rates model. Under the independent-rates model, a gamma prior is assigned on the variance of the logarithm of the rate: $\sigma^2 \sim G(1, 8)$: here again $\alpha = 1$ is chosen to represent a diffuse prior while the mean (1/8) is chosen as the reciprocal of prior mean of the root age, following the recommendation of the Multidivtime program (Thorne et al., 1998; Thorne and Kishino, 2002; Rutschmann, 2005). Under the correlated-rates model, the rate of the current branch depends on the rate of the ancestral branch, with the variance of the log rate to be $t\sigma^2$, if the current branch is t time units later than the ancestral branch.

We refer to the above settings with the independent-rates model as the standard condition. We vary the prior or settings to evaluate the robustness of the analysis. In such a case, only one factor was altered at a time.

The MCMC was run for 300,000 iterations, with samples taken every five iterations, after a burnin of 10,000 iterations. At least two runs were launched to confirm the consistency between runs.

3. Results

3.1. Phylogenetic analysis

The ML and Bayesian trees are shown in Suppl. Figs. 1 and 2 respectively. The trees are unrooted, but the root is placed within monothalamiids to respect their traditional basal position (Pawlowski et al., 1999; Pawlowski and Holzmann, 2002; Ertan et al., 2004; Longet and Pawlowski, 2007) and the independent origins of two polythalamous clades (Pawlowski and Holzmann, 2002). The first polythalamous clade is composed of two orders: Rotaliida and Textulariida (see Fig. 1 and Suppl. Fig. 1). The monophyly of Rotaliida is found in both ML (Suppl. Fig. 1A and B) and Bayesian trees (Suppl. Fig. 1C and D), and with both the concatenated (Suppl. Fig. 1A and C) and partitioned data sets (Suppl. Fig. 1B and D), although the support in the PhyML (concatenated data set), RAxML (partitioned data set) and PhyloBayes (concatenated data set) trees is low, possibly due to particularly high evolutionary rates of *Ammonia* and *E. williamsoni* and low information content in the data. However, the posterior probability (PP) is high (1.00) in the MrBayes tree (partitioned data set). The monophyly of Textulariida is found in all trees and is consistent with the traditional morphology-based systematics (Loeblich and Tappan, 1987) and is thus retained in the input topology for MCMCTREE. The second polythalamous clade is composed of calcareous order Miliolida as well as some agglutinated genera (*Ammodiscus*, *Miliammina*). In all trees, *Miliammina* branches between *Ammodiscus* and *Bathysiphon*. This position is in disagreement with recent studies suggesting that this genus should be included in order Miliolida based on molecular, immunochemical and morphological features (Fahrni et al., 1997; Habura et al., 2006). The unconventional relationship may be the result of long-branch attraction since both *Ammodiscus* and Miliolida have long branches.

Therefore, this part of the tree was modified in the input tree for MCMCTREE, in order to respect palaeontological data concerning the timing of appearance in the fossil record of *Ammodiscus* (500 Ma) (node 9 in Fig. 1), *Cornuspira* (350 Ma) (node 10) and *Miliammina* (250 Ma) (node 11). It is worth mentioning that whatever is the branching order at the base of Miliolida, our phylogenetic analyses show consequently *Bathysiphon* as the sister group to Miliolida + *Ammodiscus* clade (node 8). Although the support for this relationship is low in the PhyML tree (66%), the RAxML, PhyloBayes and MrBayes trees strongly support the phylogenetic position (87%, 0.97 PP and 0.99 PP respectively). Thus, we hypothesize that the coiled tubular ancestor of *Ammodiscus* and miliolids originated from rectilinear tubular ancestor of recent *Bathysiphon*.

The relationships within the orders Rotaliida and Textulariida were not resolved in our ML and Bayesian analyses. To construct the input tree for MCMCTREE, we used the branching order within Rotaliida published by Schweizer et al. (2008), which distinguished three well supported clades. Palaeontological knowledge about the timing of appearance of *Reophax* (node 20) and *Trochammina* (node 21) was used to resolve the topology within Textulariida.

The rooted tree topology used for divergence time estimation by MCMCTREE is shown in Fig. 1. A few different input tree topologies were also used to evaluate the robustness of the posterior time estimates to the tree topology. We also examined the impact of the placement of the root.

3.2. Estimation of divergence times

3.2.1. Likelihood ratio test of the molecular clock

We conducted the likelihood ratio test of the molecular clock hypothesis (Felsenstein, 1981) on the two partitions separately. The likelihood values were conducted with the BASEML program under the clock model and the no-clock models, without fossil calibrations (Yang, 2007). The clock model estimates 33 node ages on the rooted tree and the no-clock model estimates 65 branch lengths on the unrooted tree. Twice the log likelihood difference is compared with a χ^2 distribution with 32° of freedom. For both partitions, the test rejected the molecular clock, with $p < 0.01$ for the coding-gene partition and $p < 0.001$ for the rDNA partition. The violation of the clock is also obvious from the ML and Bayesian branch lengths estimated under the no-clock model (Suppl. Fig. 1a and b).

3.2.2. Estimates of divergence times under the standard condition

Fig. 1 shows the divergence times (chronogram) obtained under the independent-rates model. The posterior time estimates were in most cases in agreement with the fossil record. The radiation of Foraminifera was estimated to have occurred during the Cryogenian period; the posterior mean of the root age was 770 Ma with the 95% credibility interval (CI) to be (650–920). Under the correlated-rates model, the estimates were 700 Ma (610–815). These are younger than previous estimates (690–1150) obtained using the SSU only under a local-clock model (Pawlowski et al., 2003). Three main reasons could explain this difference and are presented in the following. (i) At this time, only partial SSU sequences were available. (ii) Fossil calibrations were only used for major foraminiferal radiations, which reduces the quantity of fossil information incorporated in the analysis. (iii) The method employed did not allow the rate to vary freely between lineages. Indeed, the substitution rate was estimated for multilocular species only and was then applied to all branches of the tree to calculate divergence times.

Most nodes relating monothalamid species (1, 2, 5 and 6), as well as node 7 between the monothalamous *Toxisarcon alba-Shinkaiya lindsayi* and polythalamous Textulariida + Rotaliida clade, were placed before the Early Cambrian. Those estimates support the classical idea that monothalamiids represented the vast majority of Foraminifera in the beginning of their radiation.

If the extremely fast-evolving *Ammodiscus* (see the ML and Bayesian trees in Suppl. Fig. 1) was removed, the posterior estimates of the root age became 736 Ma (600–890) under the independent-rates model, similar to estimates above. Thus removal of this species did not affect the time estimates by much.

At three major nodes, there appear to exist conflicts between the fossil calibrations and the molecular time estimates. The first concerns the origin of Rotaliida (node 23). The posterior age estimate was 190 (160–215) under the independent-rates model, with the upper limit of the CI to be even greater than the maximum fossil bound of 200 Ma. The conflict was even greater under the correlated-rates model, with the Rotaliida root dated to 240 (200–290) (Fig. 2A). When the fossil calibration was removed, almost all nodes within the Rotaliida were estimated to be much older than 200 Ma: the age of the Rotaliida root was estimated to be 380 Ma (300–460).

The second conflict concerns the origin of Miliolida (node 10). The posterior estimate, at 490 Ma (410–580), was much older than the fossil minimum bound of 350 Ma, implying a huge gap between the molecular date and the earliest fossil miliolids found. The estimate under the correlated-rates model was similar. Finally, the third conflict concerns calibration nodes in Textulariida. The posterior estimate for the origin (node 20 in Fig. 1) was 430 Ma (390–490), with the lower limit of the CI to be slightly younger than the minimum fossil bound of 400 Ma. When this fossil calibration was removed, the estimate became 396 Ma (352–470). Similarly the posterior estimates for the age of node 21 within textulariids, at which a minimum bound of 350 Ma was placed, were 380 (345–430), with the lower limit of the CI slightly lower than 350 Ma. When this fossil calibration was removed, the

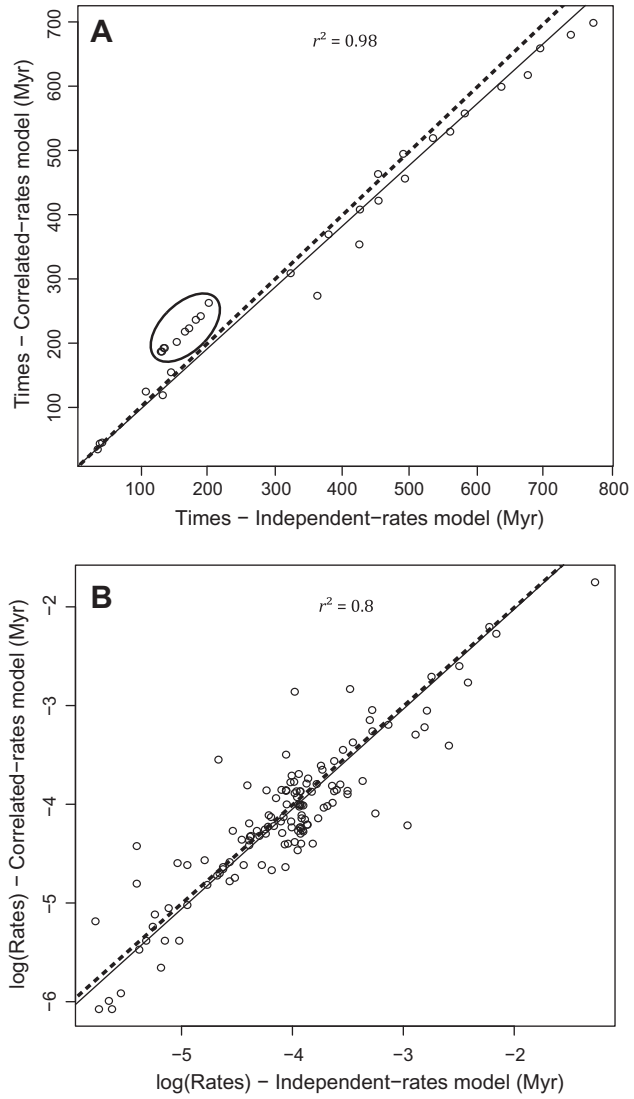

Fig. 2. Posterior means of times (A) and rates (B) estimated under the correlated-rates model plotted against those estimated under the independent-rates model. The solid line represents the regression line. The dashed line represents the $y = x$ line. The circle represents the violation of the 200 Myr maximum bound by most of the rotaliid nodes under the correlated-rates model.

estimates became 165 Ma (60–340), much younger than the fossil bound.

3.3. The impact of prior assumptions

We varied the standard condition to examine the impact of various factors on the posterior time estimation, such as the substitution model, data partitioning, the input tree topology, and the priors on times, rates and other parameters in the model.

3.3.1. The influence of the rate-drift model

In Fig. 2 the posterior estimates of times and rates were compared between the independent-rates and correlated-rates models. Except for Rotaliida and deep nodes, the posterior time estimates were very similar between the two models. The age of the root was estimated to be 50 Myr younger with the correlated-rates model than under the independent-rates model. The rates were more different between the two analyses, but the correlation in the logarithm of the rates was quite high (with $r^2 = 0.79$).

3.3.2. The influence of the prior on times

We varied the parameters (birth rate λ, death rate μ, and sampling fraction ρ) in the birth-death process with species sampling used to specify the prior on the ages of the non-calibration nodes (Yang and Rannala, 2006). Besides the values used in the standard condition ($\lambda = \mu = 2$, $\rho = 0.1$) with a nearly flat kernel density, we also considered an L-shaped kernel ($\lambda = 1$, $\mu = 4$, $\rho = 0.1$), which produces trees with long internal branches, and an inverse L-shaped kernel ($\lambda = 4$, $\mu = 1$, $\rho = 0.0001$), which produces star-like trees with short internal branches. No difference was observed among those priors for most node ages, excepted for deep nodes (Suppl. Fig. 2A and B). The results were similar to those of Yang and Rannala (2006), who also found that parameters in the birth-death process with species sampling had little influence on the posterior time estimates.

We considered the impact of the prior on the ages of the calibration nodes. First, the influence of parameters p and c in the truncated Cauchy distribution for minimum bounds was examined. Inoue et al. (2010) analyzed three empirical datasets and found that both parameters, in particular c, had a strong influence on the posterior, with larger p and c pushing up estimates of all node ages. We used $c = 1$ in comparison with $c = 0.2$ in the standard condition (Suppl. Fig. 3A). All node ages became older in the prior. However, this effect did not persist in the posterior (Suppl. Fig. 3B). Second, we used different priors on the age of the root. In comparison with the gamma distribution $G(33.2, 4.16)$ in the standard condition, we also used joint minimum and maximum bounds $B(550, 1090)$. The results are presented in Suppl. Fig. 4. This prior produced very similar estimates for the root age to the gamma prior.

3.3.3. The influence of the prior on rates

The effect of the gamma prior on the overall rate μ was examined. In the standard condition, $\mu \sim G(1, 30)$. We used $\mu \sim G(2, 60)$, so that the shape of the density is modified without changing the mean. This change in the prior produced very similar posterior estimates of times and rates (Suppl. Fig. 5).

In the standard condition, the rate-drift parameter $\sigma^2 \sim G(1, 8)$. We used $G(0.5, 8)$ and $G(10, 8)$ as well. Note that multiplying the shape parameter α by 0.5 reduces both the mean and variance by a half so that the rates are more homogeneous among lineages, and multiply α by 10 increases the mean and variance. The estimates (with those for $\alpha = 0.5$ shown in Suppl. Fig. 6) were very similar to those under the standard condition. Thus the prior assumption about the violation of the clock had little impact on the posterior time estimates. This result is in contrast with a previous study, which showed some impact of the prior on σ^2 (Inoue et al., 2010).

3.3.4. The influence of the substitution model

We used the simple JC69 substitution model for comparison with HKY85 + Γ_5 assumed in the standard condition. The time estimates were very similar between the two models (Suppl. Fig. 7). This result emphasizes that the underestimation of sequence distances produced by too simple models is balanced by the multiple fossil calibrations used in the analysis, as it was previously noticed (Yang and Rannala, 2006). However, JC69 inferred much lower rates than HKY85 + Γ_5 (Suppl. Fig. 7B).

3.3.5. The influence of the time unit used in the calculation

The birth-death process prior on times and the log-normal distribution of rates are not invariant to the change of the time unit. In theory, the results may differ when one changes the time unit from 100 Myr to 10 or 1000 Myr. We used all those three time units and found that the posterior time estimates were indistinguishable (Suppl. Fig. 8). We also reached the same conclusion using two

other datasets: the primate dataset used by Yang and Rannala (2006) and the fish Fish dataset of Inoue et al. (2010).

3.3.6. Alternative models of time and rate priors

We implemented two variations to the models of Yang and Rannala (2006) and Rannala and Yang (2007). First we replaced the birth-death kernel of Yang and Rannala (2006: equation 4) with a beta distribution when specifying the prior on times; in other words, the ages of the other nodes given the age of the root are order statistics of variables drawn from the beta distribution. Second, we implemented the gamma distribution of substitution rates in place of the log-normal under the independent-rates model. We found that the posterior estimates of times and rates were very similar to those obtained under the standard condition using the birth-death kernel for the prior on times and the log-normal distribution for the prior on rates (results not shown). The results suggest that the distributional details do not matter to the posterior time estimates in such modeling.

3.3.7. The influence of the root placement and the tree topology

Even if the root is likely to be located among monothalamids, its exact position is unknown because of the lack of closely related sister group to Foraminifera to be used as outgroups. In the topology used here, the position was somewhat arbitrarily chosen within basal monothalamids. Here we investigated the effect of the position of the root on the posterior estimates, by placing the root between all basal monothalamid species and the rest of the foraminifers. This change had little influence for almost all comparable nodes in the trees (Suppl. Fig. 9). Three nodes were estimated to be younger with the new root, but they are among the basal monothalamids, in the part of the tree where no fossil calibrations are available and where the times were hard to estimate. Their CIs were very wide and they overlap considerably between the analyses. Furthermore, the root age was estimated to be 700 Ma (600–842), compared with 770 Ma (650–920) with the previous root location. While the posterior mean was ~70 Myr younger, the posterior CIs overlap considerably.

Schweizer et al. (2008) recently proposed a well-supported phylogeny of rotaliids. They defined three major clades with strong statistical support. The present study focuses on the phylogeny of Foraminifera at a larger scale, so that many fast-evolving sites and insertions had to be deleted in the SSU alignment, preventing us to obtain an alignment supporting the Rotaliid topology found by Schweizer et al. (2008). To examine whether the Rotaliida topology is responsible for the violation of the 200 Ma maximum bound, we used a tree with the Rotaliid topology found by ML from the concatenated data (Suppl. Fig. 1A). The results were shown in Suppl. Fig. 10. All fossil calibrations within Rotaliida were respected by the posterior time estimates, and the Rotaliid root was dated to 200 Ma (180–225), indicating that the maximum bound is still violated. Thus the violation of the 200 Ma bound does not appear to be due to topological differences within Rotaliid.

3.3.8. The influence of data partitioning

We also analyzed the data as four partitions (the SSU rDNA vs. three coding genes), in comparison with the two-partition analysis under the standard condition. For the three protein coding genes, only the first and second codon positions were used, with 688, 658 and 802 sites respectively, and containing only 20, 16 and 8 species, respectively (see Suppl. Table 1). As with the two-partitions data set, the approximate and the exact likelihood calculations produced nearly identical results with the four-partitions data set. On the whole, posterior mean times are younger for the deepest nodes in the four-partitions analysis (Suppl. Fig. 11A). However, the CIs overlap widely between the two analyses. Finally, we also estimated divergence times using two mixed partitions (rDNA vs. amino acids). Suppl. Fig. 11B shows that posterior times are nearly identical to those obtained under the standard condition.

3.4. Inter-species variations of evolutionary rates

Suppl. Fig. 12 shows the two rategrams for the two partitions obtained under the standard condition. The rates for the coding-genes partition were quite homogeneous among all species and no important inter-species shifts were noticed (Suppl. Fig. 13A). However, 47% of sites in this partition are missing data. The rates were much more variable among lineages in the SSU partition, with monothalamiid and textulariid species having low rates and other parts of the tree, especially Miliolida, having high rates (Suppl. Fig. 12B). After the divergence of *Miliammina fusca* there was a fast rate acceleration, according to both the independent and correlated-rates models. This increase in evolutionary rates is accompanied by an increase in A + T content in miliolids (data not shown).

3.5. Conflicts between coding-genes and SSU within the Rotaliida and Miliolida

We analyzed the coding genes and the SSU data separately, to detect potential conflicts between the two sets of data and to examine which set is responsible for the violation of the 200 Ma maximum bound. Species having SSU sequences and no coding-gene sequence were removed from the SSU data set, so that only comparable nodes remained in the input topology. The posterior means of times estimated from the two datasets are shown in Fig. 3, which shows that four major discrepancies exist between coding genes and SSU. First, coding genes tend to predict younger times for the deeper nodes than the SSU. The root age estimate was 695 Ma (560, 860) from the genes and 750 Ma (580, 960) from the SSU. Interestingly, when the 200 Ma maximum bound for the origin of Rotaliids (node 23 in Fig. 1) was removed, the genes and SSU were in agreement concerning the times of deep nodes (Fig. 3B). Second, the two datasets gave very different age estimates for node 3 in Fig. 1, representing the divergence between *Crithionina delacai* and *Allogromia* + *Edaphoallogromia australica*: 260 Ma (120–520) for the genes and 610 (390–840) for the SSU. This node is far away from fossil calibrations and thus hard to date reliably. Third, the estimated ages of nodes for the most recent miliolids were older from the genes than from the SSU. For example, node 12 in Fig. 1, which groups *Quinqueloculina sp.* + *P. peruviana* with the taxon *P. pertusus* + *Sorites* + *Marginopora* + *A. hemprichii*, was estimated to be 400 Ma (170–490) from the genes and 160 Ma (100–300) from the SSU. The extremely long branch between nodes 11 and 12 (Fig. 1) for the SSU (Suppl. Fig. 12B) appears to explain why MCMCTREE infers a wide time interval between the two nodes and a young age for the most recent miliolids (node 12). Fourth, as with the most recent miliolids, coding-genes tend to predict older nodes within the Rotaliida. The origin of Rotaliida is dated at 190 Ma (160, 215) with the genes and at 160 Ma (120, 200) with SSU. Thus, SSU seemed to be in agreement with the 200 Ma maximum bound while the genes were not. When the maximum bound was removed, the estimated age for the origin of Rotaliida remained reasonable with SSU, at 230 Ma (120, 400), but questionable with the genes, at 380 Ma (300, 450) (Fig. 3B).

3.6. Infinite-sites plot

For a fixed set of fossil calibrations, the errors in the posterior time estimates will not approach zero when the amount of sequence data increases. Instead the joint posterior distribution will become one-dimensional, and as a result, the posterior CI widths will be linear with the posterior means (Yang and Rannala, 2006). One can then plot the posterior CI width against the

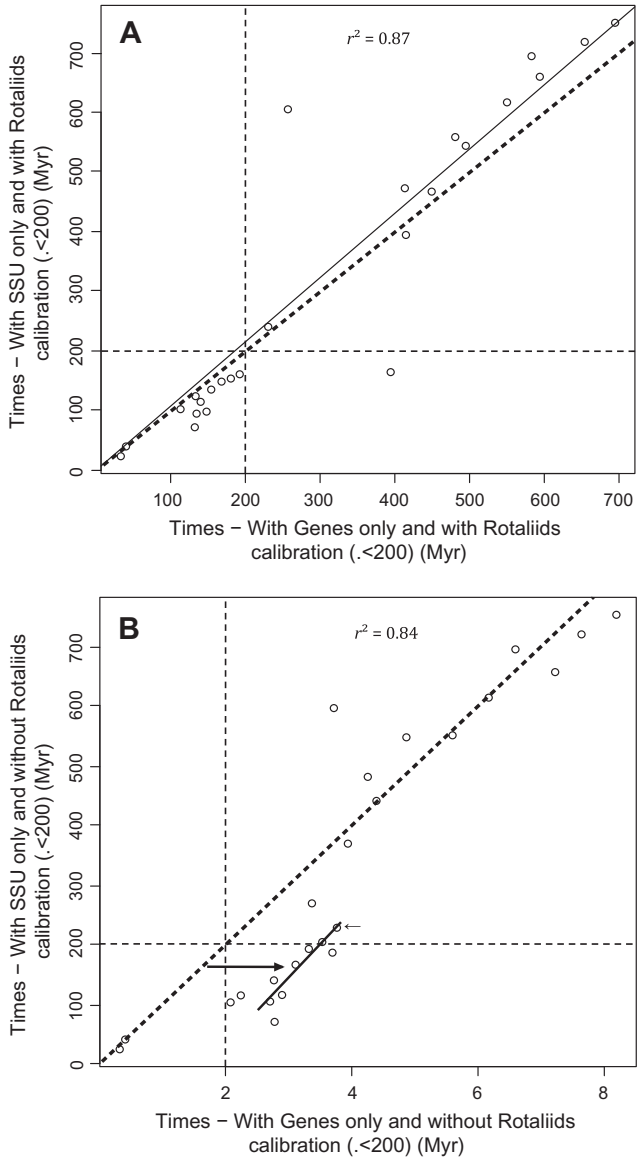

Fig. 3. Posterior means of times estimated using the protein-coding genes plotted against the posterior means of times estimated using the SSU, either with maximum bound on the root of Rotaliids used (A) or removed (B). The thin dashed lines represents the 200 Myr maximum bound on the root of rotaliids. The thick arrow represents the shift of time estimates of the rotaliid nodes.

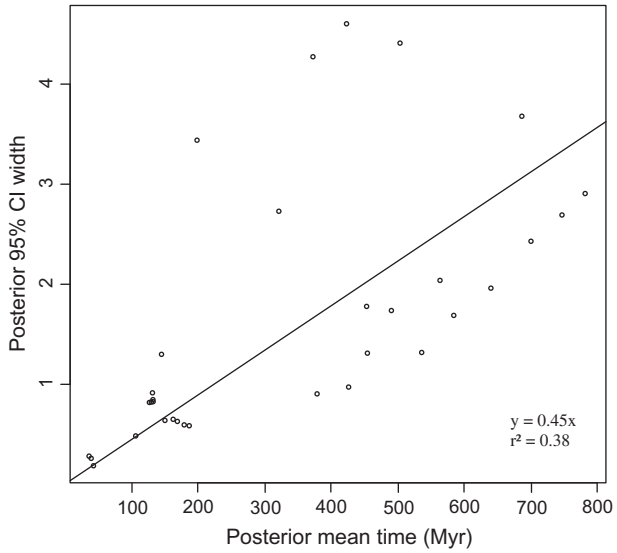

Fig. 4. The infinite-sites plot in which the widths of the posterior 95% CIs are plotted against the posterior means of divergence times. Six nodes (all above the regression line) with very wide CIs are highlighted by asterisks in Fig. 1B. Three of them belong to Monothalamiida, where no fossil calibration is available.

posterior mean of node ages to assess whether the amount of sequence data is near saturation or additional sequence data are likely to increase the precision of estimation. The infinite-sites plot for the two-partition data is shown in Fig. 4. The slope (0.45) means that every 1 Myr of divergence adds about 0.45 Myr of uncertainty in the posterior CI. The considerable scatter in the plot (with r^2 = 0.38) suggests that the sequence data are rather limited and sequencing new genes or adding species will very likely lead to more precise estimates. It is noteworthy that three of the most-poorly dated nodes are ancestors of monothalamiid species, in the part of the tree with no fossil calibrations.

4. Discussion

While molecular data has not yet completely resolved the phylogenetic relationships or produced definitive estimates of divergence times in Eukaryotes or Foraminifera, our time estimates are much more precise than in previous studies (Berney and Pawlowski, 2006; Douzery et al., 2004). This is clearly due to our use of multiple fossil calibrations and the expanded molecular dataset. Our estimates suggest that the radiation of Foraminifera occurred around 750 Ma, between approximately 650 and 900 Ma. This estimate is much younger and more precise than the estimate of Pawlowski et al. (2003), who proposed, on the basis of an SSU-only analysis, a CI of (690–1150). Our new estimate is more in line with the non-basal phylogenetic position of Foraminifera among Rhizaria in SSU phylogenies (Pawlowski and Burki, 2009). Our estimates support the hypothesis that all eukaryotic super-groups emerged during the Neo-Proterozoic (Berney and Pawlowski, 2006; Douzery et al., 2004). As the oldest fossils unequivocally attributed to Foraminifera are dated from the Early Cambrian, the results imply that the evolutionary history of this group includes a long non-fossilized period. According to our estimates, almost all divergence events among monothalamids occurred before the Cambrian explosion (Fig. 1) during the Cryogenian period (Neo-Proterozoic). Those results indicate that the Cryogenian oceanic fauna was composed of monothalamous foraminiferal species and that Foraminifera are important to the Neo-Proterozoic protistan communities and the difficulties in finding fossil traces of foraminifers during this period could be due to non-preservation of these specimens in rocks. Some microfossils, dated from the Neo-Proterozoic, were previously discovered (Porter and Knoll, 2000; Rasmussen et al., 2002). Nevertheless, it was complicated to attribute them to particular lineages because of too simple morphologies, which prompted Pawlowski et al. (2003) to suggest a reevaluation of the interpretations in these studies concerning whether these fossils represent unilocular foraminifers. The results presented here strongly support such a conclusion.

On the whole, the divergence times estimated using the molecular data are in agreement with the fossil record. The soft bounds implemented in the MCMCTREE program appear useful for detecting possible conflicts between the fossils and the molecules, as indicated by the posterior CIs going beyond the specified bounds. We noted that the 400 Ma minimum bound for the origin of Textulariids was slightly violated while the 350 Ma minimum bound within this group was likely to be inappropriate (see Section 3). This could be explained by the fact that only *Reophax* possesses gene sequences in the coding-genes partition and that the SSU

evolutionary rates for the four textulariid species are really low (Suppl. Fig. 12B). By allowing the algorithm to propose dates that do not conform with fossil calibrations during the MCMC calculation, MCMCTREE appears to have detected a more general problem concerning the uncertain status of the genus *Trochammina*. Theoretically, this is a very old genus, which appeared in Carboniferous (350 Ma) (Haynes, 1981) but its taxonomic definition is unclear and it is possible that the genus is not monophyletic. Thus we cannot rule out the possibility that the *Trochammina sp.* that was sequenced is not so old and that its divergence with *Eggerelloides scabrum* + *Textularia sagittula* occurred much later. In agreement with this, both *Textularia* and *Eggerelloides* are relatively recent genera that appeared in Paleocene and Holocene, respectively.

Furthermore, the maximum bound of 200 Ma for the radiation of Rotaliida is violated by the molecular time estimates. Under the correlated-rates model, both the mean and the CI upper limit are older than 200 Ma (see Section 3). Two reasons may explain the results. First, the fossil calibration may not represent the true evolutionary history of this group, which may have diverged from aragonitic lineage earlier in the Triassic (Haynes, 1981). Second, the dating method may not handle fast rates within Rotaliida and as a consequence overestimates the time of divergence. The SSU data set was particularly subjected to strong inter-species variation of rates within Rotaliida (Suppl. Fig. 12B), with particularly high rates for *Ammonia* and *E. williamsoni*. As it appeared that it was the coding-genes that are not in agreement with the fossil calibration, further investigations are needed to understand the impact of fast rates of coding-genes within Rotaliida on posterior time estimation. We can conclude that, although estimating the divergence time of this group is non-trivial, an origin of modern Rotaliida during the Jurassic is still plausible in respect to the SSU data set. The third conflict between molecular and fossil data concerns the origin of Miliolida. In the fossil record, miliolids appear during the Carboniferous around 350 Ma, with the first species assigned to the family Cornuspiridae (Haynes, 1981). The genus *Miliammina* emerged in the fossil record during the Triassic (from 251 to 200 Ma). Our main analysis places the radiation of Miliolida around 490 Ma, under both the independent- and correlated-rates models and with analogous results between SSU and coding-genes when these are analyzed separately. These results would imply that a period of 100–140 Myr of the miliolids history did not leave fossil traces in sedimentary rocks. Moreover, our estimates place the time of divergence between *M. fusca* and the rest of miliolids during the Devonian or Silurian between 390 and 450 Ma, depending on the data set analyzed. Thus a similar conclusion can be drawn for *Miliammina*, with a non-fossilized period of 200 Myr.

The following factors may explain why our results are not compatible with the classical scheme followed by palaeontologists. (i) Both the branches leading to *Cornuspira* and node 12 are characterized by rapid evolutionary rates in SSU (Suppl. Fig. 12B). MCMCTREE may encounter difficulties to handle those extreme high rates. (ii) The origin of Miliolida corresponds to the apparition of calcareous tests. Consequently, the early *Cornuspira* could be very weakly calcified and therefore not preserved in the fossil record. (iii) The common ancestor of Miliolida was actually still agglutinated and *Cornuspira* developed a calcified test independently from the other miliolids. However, this third hypothesis does not appear to be plausible for the reason that *Cornuspira* has a typical miliolid porcellaneous wall, and that it is not parsimonious to imagine that the formation of such wall originated more than once. Similarly, the very early divergence of *Miliammina* is unexpected since this genus has particular type of chambers winding in different planes, typical for *Quinqueloculina* and other Miliolacea that appeared in Jurassic. It is unlikely that foraminifers with such characteristic type of test have not been noticed in the fossil record by palaeontologists.

Accurately estimating divergence times during the Neo-Proterozoic among Rhizaria is necessary to understand the speciation dynamic of early Eukaryotes. In particular, inferring the time of the radiation of Radiolaria, sister-group of Foraminifera according to recent results (Burki et al., 2010), will be of great interest to understand the origin of Foraminifera. Thus, accumulation of more sequence data will most likely lead to more precise time estimates, particularly in groups of unilocular species where no fossil calibration is available, as indicated by the infinite-sites plot. Genetic data less prone to evolutionary rate variation than SSU will be especially valuable in resolving the conflicts between fossil and molecular dates observed in this study.

Acknowledgments

We are grateful to an anonymous referee for many constructive comments. The authors would like to sincerely thank Yurika Ujiie for providing some of the gene sequences used in this study. M.G. was supported by the École Normale Supérieure (ENS) of Lyon. J.P. was supported by the Swiss National Science Foundation Grant 31003A-125372. Z.Y. gratefully acknowledges the support of K.C. Wong Education Foundation, Hong Kong.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.06.008.

References

Bell, C.D., Donoghue, M.J., 2005. Dating the dipsacales: comparing models, genes, and evolutionary implications. Am. J. Bot. 92, 284–296.
Berney, C., Pawlowski, J., 2006. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proc. R. Soc. B 273, 1867–1872.
Burki, F., Kudryavtsev, A., Matz, M.V., Aglyamova, G.V., Bulman, S., Fiers, M., Keeling, P.J., Pawlowski, J., 2010. Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. BMC Evol. Biol. 10, 377.
Culver, S.J., 1991. Early cambrian Foraminifera from West Africa. Science 254, 689–691.
Douzery, E.J.P., Snell, E.A., Bapteste, E., Delsuc, F., Philippe, H., 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. USA 101, 15386–15391.
Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 5, 113.
Ertan, K.T., Hemleben, V., Hemleben, C., 2004. Molecular evolution of some selected benthic foraminifera as inferred from sequences of the small subunit ribosomal DNA. Mar. Micropaleontol. 53, 367–388.
Fahrni, J., Pawlowski, J., Richardson, S., Debenay, J.P., Zaninetti, L., 1997. Actin suggests Miliammina fusca (Brady) is related to porcellaneous rather than to agglutinated foraminifera. Micropaleontology 43, 211–214.
Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.
Gaucher, C., Sprechmann, P., 1999. Upper Vendian skeletal fauna of the Arroyo del Soldado Group, Uruguay. Beringeria 23, 55–91.
Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.
Habura, A., Goldstein, S.T., Parfrey, L.W., Bowser, S.S., 2006. Phylogeny and ultrastructure of *Miliammina fusca*: evidence for secondary loss of calcification in a miliolid foraminifer. J. Eukaryot. Microbiol. 53, 204–210.
Hasegawa, M., Thorne, J.L., Kishino, H., 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. Genes Genet. Syst. 78, 267–283.
Haynes, J.R., 1981. Foraminifera. John Wiley & Sons, New York, NY, 433 p..
Holzmann, M., Hohenegger, J., Hallock, P., Piller, W.E., Pawlowski, J., 2001. Molecular phylogeny of large miliolid foraminifera (Soritacea Ehrenberg 1839). Mar. Micropaleontol. 43, 57–74.
Holzmann, M., Habura, A., Giles, H., Bowser, S.S., Pawlowski, J., 2003. Freshwater foraminiferans revealed by analysis of environmental DNA samples. J. Eukaryot. Microbiol. 50, 135–139.
Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Inoue, J., Donoghue, P.C.J., Yang, Z., 2010. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. Syst. Biol. 59, 74–89.
Keeling, P.J., Burger, G., Durnford, D.G., Lang, B.F., Lee, R.W., Pearlman, R.E., Roger, A.J., Gray, M.W., 2005. The tree of eukaryotes. Trends Ecol. Evol. 20, 670–676.
Kostka, M., Uzlikova, M., Cepicka, I., Flegr, J., 2008. SlowFaster, a user-friendly program for slow-fast analysis and its application on phylogeny of Blastocystis. BMC Bioinform. 9, 341.
Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288.
Lejzerowicz, F., Pawlowski, J., Fraissinet-Tachet, L., Marmeisse, R., 2010. Molecular evidence for widespread occurrence of Foraminifera in soils. Environ. Microbiol. 12, 2518–2526.
Loeblich Jr., A.R., Tappan, H., 1987. Foraminiferal Genera and Their Classification. Van Nostrand Reinhold Company, New York, 2047 pp.
Longet, D., Pawlowski, J., 2007. Higher-level phylogeny of Foraminifera inferred from the RNA polymerase II (RPB1) gene. Eur. J. Protistol. 43, 171–177.
McIlroy, D., Green, O.R., Brasier, M.D., 2001. Paleobiology and evolution of the earliest agglutinated Foraminifera: platysolenites, spirosolenites and related forms. Lethaia 34, 13–29.
Pawlowski, J., Bolivar, I., Guiard-Maffia, J., Gouy, M., 1994. Phylogenetic position of Foraminifera inferred from LSU rRNA gene sequences. Mol. Biol. Evol. 11, 929–938.
Pawlowski, J., Bolivar, I., Fahrni, J.F., de Vargas, C., Gouy, M., Zaninetti, L., 1997. Extreme differences in rates of molecular evolution of Foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. Mol. Biol. Evol. 14, 498–505.
Pawlowski, J., Bolivar, I., Fahrni, J.F., de Vargas, C., Bowser, S.S., 1999. Molecular evidence that Reticulomyxa filosa is a freshwater naked foraminifer. J. Eukaryot. Microbiol. 46, 612–617.
Pawlowski, J., Holzmann, M., 2002. Molecular phylogeny of Foraminifera – a review. Eur. J. Protistol. 38, 1–10.
Pawlowski, J., Holzmann, M., Berney, C., Fahrni, J., Gooday, A.J., Cedhagen, T., Habura, A., Bowser, S.S., 2003. The evolution of early Foraminifera. Proc. Natl. Acad. Sci. USA 100, 11491–11498.
Pawlowski, J., Burki, F., 2009. Untangling the phylogeny of amoeboid protists. J. Eukaryot. Microbiol. 56, 16–25.
Porter, S.M., Knoll, A.H., 2000. Testate amoebae in the Neoproterozoic era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. Paleobiology 26, 360–385.
Rannala, B., Yang, Z., 2007. Inferring speciation times under an episodic molecular clock. Syst. Biol. 56, 453–466.
Rasmussen, B., Bengtson, S., Fletcher, I.R., McNaughton, N.J., 2002. Discoidal impressions and trace-like fossils more than 1200 Million years old. Science 296, 1112–1115.
Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Rutschmann, F., 2005. Bayesian molecular dating using PAML/multidivtime. A step-by-step manual. University of Zurich, Switzerland. <http://www.plant.ch>.
Schweizer, M., Pawlowski, J., Kouwenhoven, T.J., Guiard, J., van der Zwaan, B., 2008. Molecular phylogeny of Rotaliida (Foraminifera) based on complete small subunit rDNA sequences. Mar. Micropaleontol. 66, 233–246.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Thorne, J.L., Kishino, H., Painter, I.S., 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15, 1647–1657.
Thorne, J.L., Kishino, H., 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51, 689–702.
Ujiié, Y., Kimoto, K., Pawlowski, J., 2008. Molecular evidence for an independent origin of modern triserial planktonic foraminifera from benthic ancestors. Mar. Micropaleontol. 69, 334–340.
Yang, Z., 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42, 587–596.
Yang, Z., 2006. Computational Molecular Evolution. Oxford University Press, Oxford, England.
Yang, Z., Rannala, B., 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23, 212–226.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.